Facoltà di Sociologia - A.A. 2005-2006
Esame di Analisi Multivariata – Prof.ssa Mecatti
Ultimo recupero
12.01.2007

Avvertenza: Fornire le formule utilizzate e tutti i passaggi dei calcoli eseguiti.
Utilizzare almeno 2 cifre decimali.

Esercizio 1

a) È noto, sulla base di alcuni studi sul mercato del lavoro italiano, che il 15% dei lavoratori dipendenti italiani non risulta avere un contratto regolare.
(i) Calcolare la probabilità che in un campione bernoulliano di 7 lavoratori del mercato italiano ve ne siano 2 con un contratto non in regola.
(ii) Se il campione fosse stato di numerosità pari a 200, quanti lavoratori con un contratto non in regola ci saremmo potuti aspettare?

b) Da alcuni recenti studi risulta che la spesa media settimanale della famiglia media italiana in prodotti acquistati nei centri commerciali risulta in costante aumento. Nel 2000, a 500 capifamiglia è stato chiesto quale fosse la spesa media settimanale familiare in prodotti acquistati nei centri commerciali: tale spesa è risultata essere distribuita come una v.c. normale di media 50 euro e s.q.m. pari a 15 euro. Nel 2006, agli stessi capifamiglia è stata posta la stessa domanda. È risultata ancora una spesa distribuita come una v.c. normale ma con media 70 euro e s.q.m. 30 euro.
(i) Calcolare per l'anno 2000 e per l'anno 2006 la probabilità che una famiglia scelta a caso spenda più di 80 euro alla settimana.
(ii) Calcolare per l'anno 2000 e per l'anno 2006 la probabilità che una famiglia scelta a caso spenda meno di 50 euro alla settimana.
Confrontare e commentare i risultati ottenuti.

c) Da una popolazione sulla quale un fenomeno *X* ha media μ ignota, si estrae un campione casuale di ampiezza *n*. Allo scopo di stimare μ si considerano i seguenti stimatori:

$$T_1 = \frac{n^2 + 2n + 4}{n^3 - 8} \sum_{i=1}^{n} X_i \;,\; T_2 = \frac{n^2 - 2n + 4}{n^3 + 8} \sum_{i=1}^{n} X_i \;.$$

Esporre e discutere la proprietà di non distorsione e verificare, fornendo tutti i passaggi, se i due stimatori sono non distorti per μ e, in caso contrario, indicare una procedura per correggerli.

Esercizio 2

a) Da un'indagine italiana sulle preferenze dei lavoratori dipendenti in merito alla scelta tra pensione statale e pensione integrativa privata, condotta su un campione bernoulliano di 175 lavoratori, è risultato che 112 di essi preferirebbero avere una pensione integrativa privata. Specificando le ipotesi teoriche adottate, costruire un intervallo di confidenza al 95% per la *percentuale* di lavoratori italiani che darebbero la preferenza a una pensione integrativa privata.

b) Con riferimento al punto precedente, quale deve essere la numerosità campionaria minima per avere un intervallo di confidenza a livello 95% ma di ampiezza non superiore a 22 punti percentuali?

c) Discutere comparativamente i vantaggi e gli svantaggi di una stima intervallare rispetto ad una stima puntuale.

Esercizio 3

a) Un'associazione di istituti di credito sostiene che la percentuale dei clienti "in sofferenza", che cioè hanno o hanno avuto problemi di pagamento con le banche, si attesta intorno al 10%. Da un campione di 200 clienti di banche risulta invece che 24 di essi hanno problemi di pagamento. Verificare con un test statistico (bilaterale) al livello dell'1% di significatività, se l'affermazione dell'associazione degli istituti di credito è supportata dall'evidenza campionaria (*Suggerimento*: usare la statistica test $\frac{\left|\hat{P} - p_0\right|}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, dove $\hat{P}$ è la proporzione campionaria, p_0 quella per la popolazione ipotizzata nell'ipotesi nulla e *n* è la numerosità campionaria).

b) Col proposito di investigare sulle pari opportunità tra sessi nelle carriere lavorative è stata condotta un'indagine sui lavoratori di un grande supermercato italiano. Su 180 lavoratori, 120 sono risultati femmine e 60 maschi. Le qualifiche (opportunamente raggruppate) sono risultate le seguenti (ordinate in base alla remunerazione in senso decrescente):
1) Responsabile (di reparto, degli acquisti, del magazzino, ecc.);
2) Impiegato amministrativo;
3) Operatore al banco;
4) Operatore alle casse e agli scaffali.
L'indagine ha prodotto i seguenti risultati:

Y / ***X***	**Maschi**	**Femmine**	**Totale**
Responsabili	20	15	**35**
Impiegati amministrativi	20	30	**50**
Operatori al banco	10	35	**45**
Operatori alle casse e agli scaffali	10	40	**50**
Totale	**60**	**120**	**180**

Verificare a livello $\alpha = 0,05$, se esiste indipendenza tra *X* e *Y* utilizzando un opportuno test statistico. Interpretare il risultato statistico con riferimento alle pari opportunità tra i lavoratori del supermercato.

c) Esporre il concetto e l'utilità del *p*-value. Darne una interpretazione per l'esecuzione di test statistici a 1 o a 2 code.

Facoltà di Sociologia - A.A. 2005-2006
Esame di Analisi Multivariata – Prof.ssa Mecatti
Soluzioni appello del 12.01.2007[1]

Esercizio 1.

a) La singola prova consiste nell'"estrazione" casuale di un lavoratore italiano che può avere un contratto regolare oppure lavorare "in nero". Tale prova viene ripetuta per *n*=7 volte. La probabilità di estrarre un lavoratore "in nero" è data da $p = 0{,}15$. Abbiamo quindi tutti gli elementi per applicare il modello binomiale la cui legge è data dalla seguente formula:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

(i) Si deve valutare la legge binomiale con i parametri *n*=7 e *X*=2. Si ottiene facilmente:

$$P(X = 2) = \frac{7!}{2!5!}(0{,}15)^2 (0{,}85)^5 = 21 \cdot 0{,}0225 \cdot 0{,}4437053125 = 0{,}20965.$$

(ii) Si deve calcolare il valore atteso della distribuzione binomiale, data da *np*. Si ha quindi:

$$E(X) = np = 200 \cdot 0{,}15 = 30.$$

b)

(i) Il quesito si risolve esplicitando le seguenti probabilità per una variabile casuale normale *X*:

$$P_{2000}(X > 80) = P(Z > \frac{80 - 50}{15}) = P(Z > 2) = 1 - 0{,}9772 = 0{,}0228;$$

$$P_{2006}(X > 80) = P(Z > \frac{80 - 70}{30}) = P(Z > 0{,}333) \cong 1 - 0{,}6293 = 0{,}3707.$$

(ii) Allo stesso modo si trova:

$$P_{2000}(X < 50) = P(Z < \frac{50 - 50}{15}) = P(Z < 0) = 0{,}5;$$

$$P_{2006}(X < 50) = P(Z < \frac{50 - 70}{30}) = P(Z < -0{,}666) = P(Z > 0{,}666) = 1 - P(Z < 0{,}666) \cong$$
$$\cong 1 - 0{,}7454 = 0{,}2546.$$

c)

(i) Lo stimatore *T* per il parametro θ si dice non distorto se $E[T] = \theta$.

---

[1] A cura di Giancarlo Manzi

$$E[T_1]=E\left[\frac{n^2+2n+4}{n^3-8}\sum_{i=1}^{n}X_i\right]=\frac{n^2+2n+4}{n^3-8}E\left[\sum_{i=1}^{n}X_i\right]=i.i.d.=\frac{n^2+2n+4}{(n-2)(n^2+2n+4)}n\mu=$$

$$=\frac{n}{n-2}\mu$$

Si conclude che T_1 NON è stimatore non distorto per μ.

$$E[T_2]=E\left[\frac{n^2-2n+4}{n^3+8}\sum_{i=1}^{n}X_i\right]=\frac{n^2-2n+4}{n^3+8}E\left[\sum_{i=1}^{n}X_i\right]=i.i.d.=\frac{n^2-2n+4}{(n+2)(n^2-2n+4)}n\mu=$$

$$=\frac{n}{n+2}\mu$$

Si conclude che T_2 NON è stimatore non distorto per μ.

(ii) Presi $\tilde{\tilde{T}}_1=\frac{n-2}{n}T_1$ e $\tilde{\tilde{T}}_2=\frac{n+2}{n}T_2$ si ottiene:

$$E\left[\tilde{\tilde{T}}_1\right]=E\left[\frac{n-2}{n}T_1\right]=\frac{n-2}{n}E[T_1]=\frac{n-2}{n}\frac{n}{n-2}\mu=\mu\,,$$

$$E\left[\tilde{\tilde{T}}_2\right]=E\left[\frac{n+2}{n}T_1\right]=\frac{n+2}{n}E[T_2]=\frac{n+2}{n}\frac{n}{n+2}\mu=\mu\,.$$

Esercizio 2.

a) La percentuale campionaria è pari a:

$p_{\%}=\frac{112}{175}\times 100=64\%$.

Il livello di confidenza è:

$\alpha=0{,}05$.

La numerosità campionaria (n=175) è sufficientemente elevata; possiamo allora costruire l'intervallo di confidenza asintotico all'$(1-\alpha)\times 100$ per la percentuale di tutti i lavoratori che preferirebbero avere una pensione integrativa privata:

$$P(p_{\%}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_{\%}(1-p_{\%})}{n}}\leq P_{\%}\leq p_{\%}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_{\%}(1-p_{\%})}{n}})=0{,}95.$$

L'intervallo di confidenza sarà dunque:

$(64-z_{0,975}\sqrt{\frac{64\times 36}{175}};64+z_{0,975}\sqrt{\frac{64\times 36}{175}})$, e quindi:

$(64-1{,}96\sqrt{\frac{64\times 36}{175}};64+1{,}96\sqrt{\frac{64\times 36}{175}})$.

Svolgendo i calcoli, si trova che l'intervallo di confidenza richiesto è dato da $64\pm 7{,}11$.

b)Deve essere:

$2z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq 0{,}22$, cioè:

$2\times 1{,}96\sqrt{\frac{0{,}64(0{,}36)}{n}} \leq 0{,}22$, ed elevando al quadrato ambo i membri:

$4\times 3{,}8416\frac{0{,}2304}{n} \leq 0{,}0484$, ed infine esplicitando n:

$n \geq \frac{15{,}3664\times 0{,}2304}{0{,}0484} = 73{,}15 \Rightarrow 74$.

c) Si vedano appunti e/o libri di testo.

Esercizio 3.

a)

Il sistema d'ipotesi (bilaterale) è dunque:

$H_0 : p_0 = 0{,}10$

$H_1 : p_0 \neq 0{,}10$

Si ha:

n=200;

$\hat{P} = \frac{24}{200} = 0{,}12$;

$\alpha = 0{,}01$.

Si può utilizzare il test Z normale perché la numerosità campionaria è sufficientemente elevata. Allora, la statistica-test sarà la seguente:

$\frac{\left|\hat{P} - p_0\right|}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, che assume nel campione il valore:

$$\frac{\left|0{,}12 - 0{,}10\right|}{\sqrt{\frac{0{,}10(0{,}90)}{200}}} = \frac{0{,}02}{0{,}0212} = 0{,}943.$$

Poiché il valore critico è $z_{0,975} = 2{,}58 > 0{,}943$ NON si può rifiutare l'ipotesi nulla: si conclude che l'associazione ha fornito dati verosimili.

b)

(i) Si deve utilizzare il test d'indipendenza secondo cui si rifiuta l'ipotesi nulla H_0 di indipendenza se $\chi^2 > \chi^2_{1-\alpha;(h-1)(k-1)}$

Il valore empirico del chi quadrato è dato da:

$\chi^2 = 16{,}422$.

Poiché $\chi^2_{0,95;3} = 7{,}8147$ che è minore di 16,422, si rifiuta l'ipotesi nulla di indipendenza tra le due variabili. Le donne non sono trattate con pari opportunità nel supermercato.

c) Si vedano appunti e/o libri di testo.